Predicting Experienced Travel Time with Neural Networks: A PARAMICS Simulation Study

Charles D. Mark, Adel W. Sadek, and Donna Rizzo

Abstract— **The implementation of Intelligent Transportation Systems (ITS) in recent years has resulted in the development of systems capable of monitoring roadway conditions and disseminating traffic information to travelers in a network. However, the development of algorithms and methodologies specialized in handling large amounts data for the purpose of real-time control has lagged behind the sensing and communication technological developments in ITS. In this study data generated by a PARAMICS model of a real-world freeway section is used to develop an artificial neural network (ANN) capable of predicting experienced travel time between two points on the transportation network. Computational experiments demonstrate that the studied ANNs were able to reasonably predict experienced travel time. Generally, the study shows that the length of the time lag did not have a statistically significant effect on ANN performance, that speed appears to be the most influential input variable, and no statistically significant difference in ANN performance was observed when data from the left lane loop detector was substituted for data from the right lane loop detector.**

I. Introduction

The implementation of Intelligent Transportation Systems (ITS) in recent years has resulted in the development of systems capable of monitoring roadway conditions and disseminating traffic information to travelers in a network. It is now possible to embed sensors in a roadway to record traffic flow information in real-time and to use numerous methods ranging from radio to variable message signs (VMS) to relay traffic conditions to motorists. However, the development of algorithms and methodologies specialized in handling large amounts data for the purpose of real-time control has lagged behind the sensing and communication technological developments in ITS.

Manuscript received April 1, 2004. This work was supported by the National Science Foundation Grant CMS-0133386 and by an Eisenhower Graduate Transportation Fellowship from the Federal Highway Administration and the National Highway Institute

C.D. Mark is a Civil Engineering graduate student at the University of Vermont, Burlington, VT 05405 (e-mail: cmark@emba.uvm.edu).

A. W. Sadek , is an Associate Professor with the Department of Civil and Environmental Engineering, University of Vermont, Vermont, Burlington, VT 05405 (phone: 802-656-4126 fax 802-656-8446 e-mail: asadek@emba.uvm.edu).

D. Rizzo , is an Assistant Professor with the Department of Civil and Environmental Engineering, University of Vermont, Vermont, Burlington, VT 05405 (e-mail: drizzo@emba.uvm.edu).

The goal of Dynamic Route Guidance (DRG), one of many ITS strategies, is to recommend routes to motorists especially when unexpected capacity reducing incidents occur on a transportation network. One of the many DRG strategies, feedback control, will be the focus of this study.

Feedback control methods *feed* information from the output of a system *back* into the system controller to achieve a particular objective. In DRG applications, the output signal is typically an observed travel time across a section of the network. Numerous research efforts have focused on the development of this type of feedback control [1]-[3]. These methods feedback information pertaining to current travel times (for example by tracking a vehicle through a network segment and reporting the vehicle's travel time) to a controller, which then strives to balance travel times between multiple routes. However, the ability of this methodology to balance travel times between multiple routes degrades in the presence of congestion. During congested flow conditions the time for a vehicle exiting a network segment to traverse that segment is not necessarily going to be equal to the time required for a vehicle entering the network segment to traverse the same distance.

The goal of this study is to design an artificial neural network (ANN) that can predict the experienced travel time across a transportation network using real-time traffic condition information (such as speed and flow) from sensors embedded in the roadway. *Experienced* travel time is the time it will take a vehicle entering a network segment at time t to traverse that segment, as opposed to the *observed* or *current* travel time which is the time it took vehicles exiting the segment at time t to traverse the segment.

II. Artificial Neural Networks and Traffic Flow Prediction

A. Introduction to ANNs

ANNs are an inductive-logic-based adaptive computational technique. Originating from early computational models of the human nervous system, ANNs have become a proven tool in pattern recognition and non-linear function mapping applications. When used to map functions, the goal is to train an ANN to relate a set of input vectors to a set of desired vectors, by providing the ANN with a finite number of training input vectors with their associated desired vectors [4].

The Multi-Layer Perceptron (MLP), one of many ANN topologies, is constructed from multiple layers of non-linear processing elements (PEs) that are either fully or partially connected to the other PE layers in the network by a matrix of weighted pathways. Initially the weights are set to random values and then adjusted by a learning algorithm, such as back propagation, as successive iterations of the training data set are introduced to the network [4].

Additionally, MLP ANNs are a stochastic modeling tool, which suggests that unless the error surface is excruciatingly simple, it is not probable that two similar MLP ANN topologies initialized with different random weights and trained on the same input data sets will result in the same exact mapping function. It is uncertain that any single trained network is the best solution and therefore false conclusions could be reached by relying on the performance of any individual network to be representative of a topology or strategy. Therefore, rather than comparing individual trained networks when evaluating the effectiveness of different topologies or strategies, it would be more accurate to compare groups of trained networks sharing the same topology and parameters, but initialized with different random weights [4].

Inductive logic based computational techniques can be helpful in modeling complex systems in which the inputs and outputs to a system are known, but the complexity of the internal system processes prevent the creation of a deductive model. ANNs are also useful computational tools when the amount of data to be analyzed would increase the computational effort required by traditional statistical techniques beyond practical computational limits. In this study, a complex relationship exists between the data collected by multiple sensors and experienced travel times. Additionally, the thousands of data points collected in real time from roadway sensors would require an enormous amount of computational effort if traditional statistical methods were implemented.

B. Traffic Flow Prediction Using ANNs

Several recent investigations have studied the ability of ANNs to predict short-term traffic conditions. Dougherty et al.[5], Smith and Demetsky[6], and Dougherty and Cobbett [7] used backpropagation ANNs to forecast short-term traffic conditions, such as volumes, speed, and occupancy. Chen and Muller[8] used dynamic ANNs to predict short-term traffic flow under incident free and incident conditions. Park and Rilett[9] developed ANNs for forecasting multiple-period freeway link travel times, as well as for one-step corridor travel time forecasting[10]. Additionally, Lint et al. [11] describe an approach for freeway travel time prediction using state-space neural networks.

However, there are several distinguishing aspects of this paper from previous studies. First, most of the previous studies attempted to predict traffic conditions or travel times in the absence of capacity incidents. In this study, the authors explicitly attempt to predict *experienced* travel times in the presence of incidents. Second, this paper includes a comprehensive statistical analysis of the impact of several design parameters on the quality of the ANN predictions. Based upon the results from this statistical analysis, conclusions are developed as to the type and presentation format of data to an ANN to achieve reliable predictions of experienced travel time.

In a previous study by Mark and Sadek [12], the ANNs were trained with a data set that mimicked the continuous output from the roadway sensors. This simulated data set consisted of 5 minute averages of the input parameters from 5:30AM until 8PM for an 85 weekday period. While the macroscopic computer model used in that study was programmed to simulate capacity reducing incidents occurring at much higher frequencies than reality, still an overwhelming proportion of the data corresponded to non-incident conditions. During these time periods, flow on the network was steady state and the travel time was simple to predict. Since the training data set consisted of relatively few instances in which incident conditions were present, the gradient decent algorithm used to train *the ANN could still achieve low error values for the data set as a whole even though the spikes in travel time representing incident conditions were not fit well.* To accommodate for this disproportion of data, specialized ANNs had to be trained for incident and non-incident conditions. *In doing so, the specialized ANN trained to fit the peaks in travel time was only able to predict with a percent error of 23.80%.*

The percentage of data representing incident conditions in the previous study by Mark and Sadek was about 7%. The current study has increased this percentage to 20%, which has allowed the training of one ANN, avoiding the need to split the data and train specialized ANNs. The ANNs trained in this study were, unlike the trained ANNs in the previous study, able to reliably predict the peaks in experienced travel time, as is discussed later. This study also uses an improved data set generated by the microscopic traffic simulation program PARAMICS, which resulted in a more realistic data set than the data set created by the macroscopic model in the previous study.

III. Experimental Design

A. Modeling the Test Transportation Network in PARAMICS

PARAMICS (PARAllel MICroscopic Simulation) was used to create a computer model of a test tranportation network used to generate the data for ANN training, cross validation, and testing. PARAMICS is a three-dimensional, microscopic traffic simulation suite consisting of 5 modules: Modeler, Analyzer, Processor, Programmer, and Estimator.

An approximately 10 mile section of I-89 near Burlington, VT was modeled in PARAMICS (Figure 1). This freeway section was comprised of 2 lanes in each direction, speed limits varying between 55 and 65 miles per hour, and 5 exits (2 full interchanges, 2 half interchanges, and 1 clover-leaf). Ten loop detectors, programmed to record flow and speed data, were evenly position approximately 1 mile apart in both the right and left lanes.

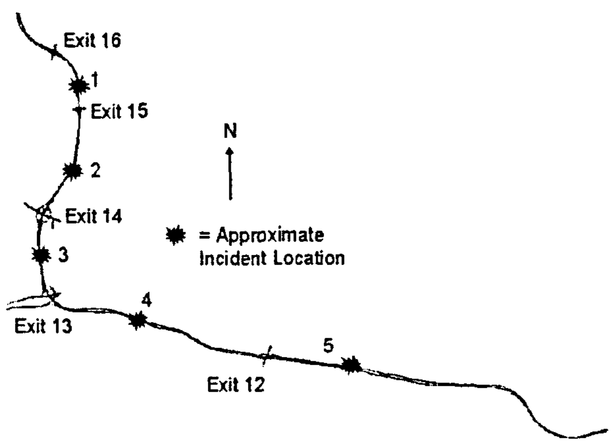

Fig 1. The Modeled Transportation Network

Aerial photo overlays were used in PARAMICS modeler to create a geometrically accurate representation of I-89. As opposed to pervious generations of microscopic traffic simulations that use generic road geometries, PARAMICS is able to model the actual road geometry, resulting in a more realistic simulation that is capable of capturing driver behavior associated with the unique features of a specific roadway. However, the added realism also complicates the task of calibrating the model, since geometry parameters must be adjusted to achieve realistic flow.

In our experience, the two primary parameters that had to be adjusted to achieve realistic flow patterns were the stop lines and the kerb points (PARAMICS was developed in Scotland, hence the British spelling of curb). The stop lines, which are specified at the beginning and end of each link, dictate the position and angle at which vehicles enter and leave a link. The kerb points define the physical boundary of the road. When the simulation was run using the default settings for these parameters, vehicles had difficulty in making a smooth transition between links which was indicated by severe reductions in speed at the junction between freeway links. Both the kerb points and stop lines were manually adjusted until vehicles could freely pass between links without a sharp reduction in speed.

PARAMICS requires an Origin-Destination (O/D) matrix to generate and assign trips on the network. A sub-matrix, corresponding to the test network, was extracted using VIPER and TP-Plus from the larger transportation planning model matrix developed for Chittenden County, VT. The O/D matrix for the test transportation network consisted of 12 zones. The two most important zones are the zones that define the external boundaries of the transportation network at the northern and southern most extremities, zones 1 and 2 respectively. *In this study, travel time is predicted for only the south-bound traffic movement on I-89 from zone 1 to zone 2.* The same methodology outlined in this paper could be employed for predicting north-bound experienced travel time, except the ANN would have to be trained using north-bound traffic data.

To create a training data set that would result in an ANN capable of generalization, the flow on the network was varied. Five different levels of flow, corresponding to roughly LOS A through LOS E, were simulated on the southbound direction of I-89. The level of flow was controlled by adjusting the hourly demand in the O/D matrix between zones 1 and 2. Additionally, merging and diverging flow occurring at the on and off ramps were included in the model to make the problem more realistic and in doing so, added more complexity to the overall traffic flow dynamics and therefore increased the difficulty in predicting experienced travel time.

In this study all incidents occurred in the right lane. Four categories of incidents were programmed in PARAMICS. Type 1 incidents lasted for approximately 10 minutes and the duration of incidents increased by 10 minute intervals up to type 4 which lasted for 40 minutes. Five incident locations were selected as depicted in Figure 1. Incident locations were selected as not to be with-in close proximity of off ramps. Evidently, through preliminary experimentation, erratic behavior was observed when incidents were placed within the influence area of an off ramp. The authors attribute this erratic behavior to the lack of an automatic diffusion feature in the PARAMICS model [13].

B. Producing and Post-Processing of Data from PARAMICS

After the PARAMICS model of I-89 was calibrated, the Processor Module, which is used to execute batches of simulations quickly, was used to run all combinations of incident types, incident locations, and flow rates for a total of 100 different scenarios. Each scenario was run for 4 hours of simulation time. During the first 30 minutes the simulation was allowed to reach steady state conditions, at 30 minutes the incident occurred, and the remainder of the 4 hours was used to recover from the incident and allow the network to return to steady state.

Point values for speed and flow were recorded by each of the loop detectors every time a vehicle passed. Additionally, the travel time for each vehicle between zone 1 (the northern extremity of the network) and zone 2 (the southern extremity of the network) was recorded. When an incident occurred in the simulation, the time, location, and type of incident was recorded.

After the 100 scenarios were simulated, the data was post-processed in MATLAB. The speed and flow data from the sensors, as well as the travel time data, were aggregated and averaged over 5-minute intervals of simulation time. Additionally a vector was created to capture incident information. This vector used discrete symbols to describe the incident situation on the network corresponding with the 5-minute intervals previously described. The absence of an incident on the transportation network at a given time was represented as a 0 value in the vector, while the presence of an incident was represented by the incident's categorical type and location as described above.

C. *ANN Development*

The MATLAB ANN toolbox was used to develop the multi-layer perceptron (MLP) topology used in this study. The MLP was constructed using one hidden layer with 8 nodes and used the hyperbolic tangent as the transfer function in both the hidden and output layers. The structure of the MLP ANN was determined though preliminary experimentation.

The nodes comprising the input layer enabled a time structure to be incorporated in the ANN, by means of a time-lag in the input data set. As illustrated in figure 2, node 1 would represent speed data from sensor 1 at time t, while node 2 would represent speed data from sensor 1 at time t-1. Input data pertaining to speed and flow collected by the loop detectors was normalized between 1 and –1. Incident data, being discrete symbolic data, was represented by a set of binary input nodes. For example, incident type data was represented by a set of 4 nodes. If no incident was present at time t, the input values to the nodes was [0 0 0 0], but if, incident type 3 was present at time t, the input values to the nodes would be represented as [0 0 1 0]. A similar method was used to represent incident location.

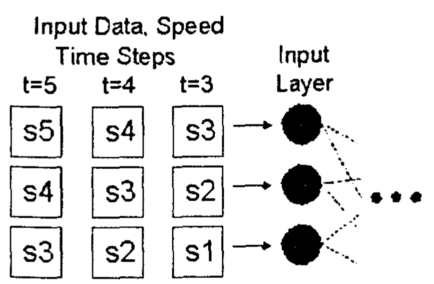

Fig. 2. Schematic illustrating how nodes comprising the input layer incorporate time-lagged variables.

The ANNs were trained using a gradient decent backpropagation algorithm, which adapted the momentum constant and the search step size as training progressed. Cross validation was used to prevent over-training of the ANN. The network trained for 1000 iterations or until the error produced by the cross validation set exhibited a significant increase. Both the cross validation and testing data sets were created using the same method used to create the training data sets, however perturbations in incident duration, incident location, and traffic volume were made to ensure that the cross validation and testing data sets were not replicas of the training data set.

IV. Computational Experiments

A. *Effect of Varying the Time Lag*

The objective of this experiment is to investigate the effect of varying the time-lag incorporated into the input data on ANN performance. In this experiment the input data set used speed and incident information while the time lag of the input data set was varied between 4 to 0 time steps (which corresponds to 20 to 0 minutes of real time). The last trial used 0 time steps which means that only current data from the sensors is used as input to the ANN.

B. *Influence of Input Variables*

Three types of input variables were available to predict travel time: speed, flow, and incident data. The incident data was comprised of two discrete classification variables, the first describing the type (i.e. duration) of the incident and the second describing the approximate location of the incident. The objective of this experiment is to determine how the performance of the ANN is influenced by these variables. Using a time lag of 4 time steps (20 minuets) the following 6 scenarios of input variables were tested: 1) Speed 2) Flow 3) Incident Data 4) Speed and Incident Data 5) Flow and Incident Data 6) Speed and Flow.

C. *Impact of Using Data from the Left Lane Loop Detector verses the Right Lane Loop Detector*

In this study the incidents occur in the right lane. The previous two experiments used data from the loop detector embedding in the right lane of the freeway. The purpose of this experiment is to investigate the ability of an ANN to predict travel time, if data from the left sensor is used instead of data from the right sensor. This experiment uses the same input variables as the second experiment, expect that data from the loop detector embedded in the left lane are used.

D. *Evaluating ANN Performance*

Due to the stochastic nature of training MLP ANNs, the ability of a topology defined by a set of specific parameters to map a data set is better represented by a group of similar networks trained with different random initial weights than by a single network. In accordance with this principle, for all computational experiments, 20 networks differing by only their initial random weights were trained.

The error measure used in this study was the absolute percent error averaged over the testing data set, defined as:

$$PE_i = \frac{\sum_{i=1}^{n} \frac{|P_i - D_i|}{D_i}}{N} \times 100\% \qquad (1)$$

where,
P_i = predicted output value from the ANN for time step i
D_i = desired output value for time step i
N = number of exemplars in the test data set

V. Results

Figure 3, a graph comparing actual and predicted values for experienced travel time, clearly illustrates that the ANNs implemented in this investigation were capable of reasonably predicting experienced travel time. It is worth noting, that while the predicted peak experienced travel times do not exactly match the desired experienced travel time values, many of the peaks are predicted with-in 1 minute of the desired experienced travel time, which is sufficient considering the natural variation in travel time between individual vehicles.

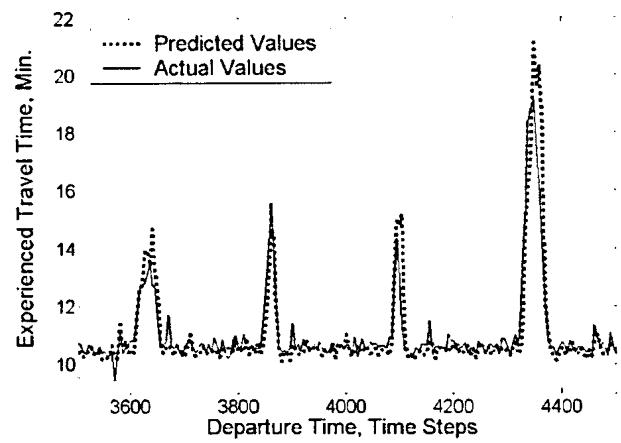

Fig. 3. Graph of departure time verses experienced travel time illustrating the ability of the ANNs studied to accurately predict experienced travel time.

A. *Effect of Varying the Time Lag*

It was initially hypothesized that increasing the time lag and therefore providing the ANN with more information pertaining to the time structure of the data would improve the ability of the ANN to predict experienced travel times. However, as seen in table 1, varying the time lag had no statistically significant impact on the ability of the ANN to predict experienced travel times. A series of F-tests and t-tests were used to determine statistical significant differences and as a result, a performance category was assigned to each trial. Performance categories with letters that appear higher in the alphabet correspond to lower levels of error. Trials that are assigned to the same performance category do not exhibit statistically significant differences in error.

This suggests that the ANNs used in this study required little or no information about the time structure of the data, which is surprising considering the inherent time structure of the data. The loop detectors were distributed in space in the computer model. Since, *shock waves move through space over time*, a temporal aspect is still incorporated in the input data set when no time lags are present for individual points in space. It should be noted that by reducing the amount of input data, the size of the ANN and therefore the computational effort needed to train the network are also reduced.

B. *Influence of Input Variables*

Table 2 clearly illustrates that the best ANN performance is achieved when speed data are used to train the network. The input data sets that used speed data alone and speed data combined with incident data both produced the lowest levels of error, 4.10% and 4.19% respectively. The input data sets that used either flow data or flow data in conjunction with incident data produced the highest levels of error, 14.98% and 15.91% respectively. The input data set using both speed and flow information produced an error of 10.60%. Surprisingly, the data set that included only the discrete binary information related to incident type and location categories resulted in a trained ANN with a percent error of 7.41%, outperforming all of the networks that incorporated flow information.

TABLE 1
EFFECT OF VARYING TIME LAG ON ANN PERFORMANCE

Time Lag, min.	Average Percent Error	Performance Category
20	4.50%	A
15	4.47%	A
10	4.18%	A
5	4.15%	A
0	4.67%	A

The reason that the ANNs trained using speed input data and speed data in conjunction with incident data exhibited the lowest percent error could be derived form the relationship between travel time and speed. Travel time and speed share a direct inverse relationship; as speed increases travel time decreases and visa versa. Incident data did not improve the ANN's estimate of experienced travel time since information about incident location and duration is indirectly implied in the speed data. However, travel time and flow are not as directly related; an increase in flow does not necessarily cause an increase in travel time. During uncongested flow conditions it is possible that an increase in flow will not cause a decrease in speed and in turn travel time will be unaffected. Traffic flow must exceed a particular level before an increase in flow will decrease speed and therefore increasing travel time. In a sense, flow is one more order removed from travel time than speed, therefore travel time is easier to predict from speed rather than flow data.

C. *Impact of Using Data from the Left Lane Loop Detector verses the Right Lane Loop Detector*

Table 3 compares the performance of ANNs trained using data from the right lane loop detectors verses the performance resulting from training ANNs using the left lane loop detectors. With the exception of the trial that used speed as the input to the ANN, there was no statistically significant difference in ANN performance between the ANNs trained using data from either the right or left lane loop detectors.

TABLE 2
INFLUENCE OF INPUT VARIABLE ON ANN PERFORMANCE

Input Data	Average Percent Error	Performance Category
Speed	4.10%	A
Flow	14.98%	D
Incident	7.41%	B
Speed, Incident	4.19%	A
Flow, Incident	15.91%	D
Speed and Flow	10.60%	C

TABLE 3
IMPACT OF USING DATA FROM THE LEFT LANE LOOP DETECTOR VERSES THE RIGHT LANE LOOP DETECTOR ON ANN PERFORMANCE

Input Data	Average Percent Error Right Loop	Average Percent Error Left Loop	Significant Difference
Speed	4.10%	3.53%	Yes
Flow	14.98%	16.56%	No
Speed, Incident	4.19%	4.10%	No
Flow, Incident	15.91%	15.38%	No
Speed and Flow	10.60%	11.67%	No

VI. CONCLUSION

This study focused on furthering the development of ANN models for predicting experienced travel times in the presence of incidents using data describing traffic conditions collected by a set of sensors distributed though-out a transportation network. A MLP ANN with time-lags incorporated into the input data set was used to relate speed, flow, and incident information to experienced travel timed. The following are the main conclusions from this study.

1) The absence of an automatic diffusion feature in the PARAMICS modeler can result in erratic behavior, especially in the proximity of off-ramps.
2) The MLP topology using time-lagged input data was capable of predicting experienced travel times in the presence of capacity reducing incidents.
3) Increasing the percentage of data corresponding to time periods influenced by incidents greatly improved the ability of a trained ANN to predicted experienced travel times when compared to previous studies.
4) Using a discrete binary representation of the categorical incident information improved the ability of a trained ANN to predict experienced travel times when compared to previous studies.
5) Varying the time lag incorporated into the input data set had no statistically significant effect on the performance of the trained ANNs.
6) The best ANN performance was achieved by using either speed data or both speed and incident data as the input data set.
7) The inclusion of flow data in the input data set resulted in the highest levels of error in ANN performance.
8) In general, no statistically significant increase or decrease in the ability of an ANN to predict experienced travel time was observed when traffic condition information recorded by the loop detector in the left lane was substituted for the same information from the right lane loop detector.

REFERENCES

[1] Jayakrishnan, R., W.K. Tsai, and A. Chen, *A Dynamic Traffic Assignment Model with Traffic-Flow Relationships.* Transportation Research, 1995. **3C**(1).

[2] Ben-Akiva, M.E., et al. Development of a Route Guidance Generation System for Real-Time Application. in IFAC Conference. 1997. Chania, Greece.

[3] Ziliaskopoulos, A.K., A Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem. Transportation Science, 2000. **34**: p. 1-12.

[4] Principe, J.C., N.W. Euliano, and W.C. Lefebvre, *Neural and Adaptive Systems: Fundamentals Through Simulations.* 2000: John Wiley and Sons, Inc.

[5] Dougherty, M., H.R. Kirby, and R.D. Boyle, *The Use of Neural Networks to Recognize and Predict Traffic Congestion.* Traffic Engineering and Control, 1993. **34**(6): p. 311-314.

[6] Smith, B.L. and M.J. Demetsky, *Short-Term Traffic Flow Prediction:Neural Network Approach.* Transportation Research Record, 1994(1453): p. 98-104.

[7] Dougherty, M. and M.R. Cobbett, *Short-Term Inter-Urban Traffic Forecasts Using Neural Networks.* International Journal of Forecasting, 1997. **13**: p. 21-31.

[8] Chen, H. and S. Grant-Muller, *Use of Sequential Learning for Short-Term Traffic Flow Forcasting.* Transportation Research Part C, 2001. 9(5): p. 319-336.

[9] Park, D. and L. Rilett, *Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks.* ASCE Journal of Transportation Engineering, 1998. **125**(6).

[10] Rilett, L. and D. Park, Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks. ASCE Journal of Transportation Engineering, 1999. **125**(6).

[11] Lint, V., S.P. Hoogendoorn, and H.J. Zuylen. Freeway Travel Time Prediction with State-Space Neural Networks. Transportation Research Board 81st Annual Meeting. 2002. Washington, D.C.

[12] Mark, C.D. and A.W. Sadek. Learning Systems for Prediciting Expereinced Travel Times in the Presence of Incidents: Insights and Lessons Learned. Transportation Research Board 83rd Annual Meeting. 2004. Washington, D.C.

[13] Yun, Ilsoo and Byungkyu Park. Development of ITS Evaluation Test-Bed Using Microscopic Simulation – City of Hampton Case Study. Center for Transportation Studies at the University of Virginia. Research Report No. UVACTS-15-0-45, 2003.